EXHIBIT 28
REDACTED

UNITED STATES DISTRICT COURT
EASTERN DISTRICT OF TEXAS
SHERMAN DIVISION

The State of Texas, et al.,	Case No. 4:20-CV-957-SDJ
Plaintiff,	
v.	
Google LLC,	
Defendant.	

REBUTTAL EXPERT REPORT OF
MICHAL A. MALKIEWICZ

HIGHLY CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

Table of Contents

HIGHLY CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

I.   INTRODUCTION & QUALIFICATIONS

1. My name is Michal A. Malkiewicz. I am a Vice President at Charles River Associates ("CRA") and an Adjunct Professor at the University of Southern California ("USC"). At USC, I teach graduate-level courses focused on using data analytics techniques in economics, finance, law, marketing, and survey research. CRA is an international consulting and expert services firm that, among other services, offers economic, financial, and strategic expertise for disputes, regulatory proceedings, policy analysis, and business strategy and planning. At CRA, I specialize in the application of economics, finance, market research, and statistics to questions arising in the provision of these services.

2. I was previously a Managing Director at Ankura, a Director at Epsilon Economics, an Associate Director at Navigant, and a Senior Associate at LECG, all of which are/were premier consulting and expert services firms. I also worked as a survey research professional at the National Opinion Research Center and as an economic research assistant to the Honorable Richard A. Posner, who at the time was a Judge on the United States Court of Appeals for the Seventh Circuit and a Senior Lecturer in Law at the University of Chicago Law School. Prior to my appointment at USC, I taught graduate-level courses in international finance and economics as a Visiting Lecturer at the Tyumen State University in Russia.

3. I received Bachelor of Arts degrees in Economics and Environmental Studies and a Master of Business Administration degree, with Honors, from the University of Chicago. With my M.B.A. degree, I was awarded concentrations in Accounting, Analytic Finance, Economics, Econometrics & Statistics, and International Business, and a Certificate in Marketing Analytics. I also hold a Master of Science degree in Applied Economics from Johns Hopkins University, where I obtained a Certificate in Forecasting Practice from the International Institute of Forecasters and completed graduate survey methodology coursework based on the curriculum of the Joint Program in Survey Methodology, a consortium graduate program founded by the U.S. Federal statistical agencies.

HIGHLY CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

4. I have studied graduate-level economics and have completed graduate-level coursework in econometrics and statistics, marketing analytics, and survey research methods, including, among others, the concepts and methods discussed in this report. I have studied, used, and taught programming languages such as Python, among others, throughout my academic and professional careers. Much of my work over the last 18 years has pertained to the application of these areas of study to questions arising in a legal context, including those arising in antitrust, breach of contract, false advertising, international trade, unfair competition, and patent, copyright, and trademark infringement disputes, among others.

5. My clients have included numerous large and small businesses, corporations, governmental entities, individuals, law firms, and non-profit entities in the United States and abroad. As both a consultant and testifying expert, I have performed analyses in hundreds of disputes and consulting engagements, including before U.S. and foreign courts, U.S. and foreign regulatory agencies, the U.S. International Trade Commission, the U.S. Federal Trade Commission, and U.S. and international arbitration organizations. I have also provided various business consulting services in the areas of economics, finance, marketing, and strategy, as well as intellectual property investing, valuation, and licensing.

6. I am a member of the American Association for Public Opinion Research, the CFA Institute, the Licensing Executives Society, and the Intellectual Property Owners Association. I am a non-lawyer member of the American Bar Association, the Chicago Bar Association, the International Council for Commercial Arbitration, and the American Intellectual Property Law Association. I have published and presented in the areas of consumer surveys, data analytics techniques, economics, and finance and served as an economic expert witness in courses offered by the National Institute for Trial Advocacy, the Kirkland Institute for Trial Advocacy, and with the Jerome Mirza Trial Academy at the University of Illinois College of Law.

7. My curriculum vitae, which details my professional qualifications, is attached as Schedule 1 to this report. CRA bills at an hourly rate of $975 for my work on this engagement. My compensation is not dependent or contingent on the outcome of this case or the substance of my opinions.

HIGHLY CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

II. ASSIGNMENT

8. On October 4, 2024, Plaintiff States submitted the Supplemental Expert Report of Prof. Jacob Hochstetler (hereafter "Hochstetler Report").[1] I have been asked by counsel for Google to respond to certain opinions contained in the Hochstetler Report in this rebuttal expert report.[2] To the extent that I do not specifically respond to any of his opinions, that should not be construed as an endorsement of any other opinions expressed in Prof. Hochstetler's Opening or Rebuttal Reports that I do not discuss here.

9. A list of the materials I relied upon in the course of preparing this report can be found in Schedule 2 to this report. My analysis may change before trial if additional information from any of the parties or their experts becomes available. I, therefore, reserve the right to supplement my report accordingly.

III. SUMMARY OF OPINIONS

10. Based on my education, training, and experience, the facts and allegations in this case, the research I conducted, my review and evaluation of the materials listed in Schedule 2 and my analysis set forth in this expert report, I have reached the following conclusions to a reasonable degree of economic and professional certainty:

   a. Prof. Hochstetler's primary opinion—"that the total number of messages not retained by Google employees subject to a litigation hold was a million and a half or more in 2022 only"[3]—is unreliable because it is based on a series of errors and unsupported assertions.

   b. Prof. Hochstetler's understanding of a "litigation hold" appears to be drawn from a publicly available glossary defining the term.[4] Google produced redacted versions of the litigation holds issued in this case to Plaintiffs.[5] Based

---

[1]   Supplemental Expert Report of Jacob Hochstetler, *The State of Texas, et. al. v. Google LLC*, United States District Court, Eastern District of Texas, Sherman Division, Case No. 4:20-CV-957-SDJ, October 4, 2024.

[2]   I have not undertaken an analysis of all chat conversations or messages collected, preserved, or produced by Google in this case because it was not necessary to do so to rebut any of Prof. Hochstetler's Opinions.

[3]   Hochstetler Report at n.4.

[4]   Hochstetler Report at n.32 (citing Thomson Reuters, "Glossary: Litigation Hold," https://uk.practicallaw.thomsonreuters.com/9-501- 9293?transitionType=Default&contextData=(sc.Default) Accessed on September 25, 2024.).

[5]   GOOG-AT-MDL-C-000087449; GOOG-AT-MDL-C-000087452; GOOG-AT-MDL-C-000087456; GOOG-AT-MDL-C-000087457.

HIGHLY CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

on the list of materials relied upon by Prof. Hochstetler,[6] it does not appear that he considered or relied upon these documents.

c. Prof. Hochstetler infers that the number of Google employees subject to a litigation hold by the beginning of 2022 was 141.[7] In fact, I understand that the Plaintiffs have obtained discovery from a much larger population of custodians.

d. Prof. Hochstetler's analysis fails to accurately reflect the litigation holds. To arrive at his conclusions, he takes the erroneous custodian count, assumes that half of these 141 employees had left the company by 2022 (with no support for this assumption), and multiplies the resulting count by the "estimated 18,566 messages lost" for five custodians over a 68 day period.[8] As explained below, these five custodians reflect what is commonly referred to as a "sample of convenience," the output of which is unreliable in this case due to highly likely biases, such as selection bias and non-response bias. Additionally, drawing any conclusions from such a small and unrepresentative sample would be improper and lacking in statistical significance.

e. Using these biased data, Prof. Hochstetler opines that the results of his analysis are "conservative" because, among other things, "one of the individuals included, Dr. Varian, had what appears uncharacteristically low messaging volume."[9] Elsewhere in his report, Prof. Hochstetler speculates without basis that "[t]his could be due to the holiday season and does not necessarily reflect [the custodian's] typical message volume."[10]

f. By Prof. Hochstetler's own calculation, this results in "1.4 million messages lost in a single year."[11] Prof. Hochstetler then adds an arbitrary 100,000 messages to that figure to arrive at his headline conclusion that the number of

---

[6]   Hochstetler Report at Appendices A and B.
[7]   Hochstetler Report at ¶¶ 8, 82, n.4, and n.106.
[8]   Hochstetler Report at n.4 and n.108.
[9]   Hochstetler Report at n.4 and n.106.
[10]   Hochstetler Report at ¶ 27 and n.53.
[11]   Hochstetler Report at n.4.

HIGHLY CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

messages lost was "a million and a half or more."[12] Moreover, Prof. Hochstetler does not account for the fact that many Google employees left the Display Ads business before the timeframe covered by his Report.[13]

g. In addition, Prof. Hochstetler's primary conclusion is stated to be based on four sub-opinions, each of which suffers from the same flaws and, in some cases, additional flaws.

h. Prof. Hochstetler opines that "at least 18,566 messages[,] out of a total of around 21,269, were lost in the 68-day period covered by the Log Dataset."[14] This conclusion is unreliable because it is based on a biased sample of convenience that is not a reliable basis from which to extrapolate conclusions about Google's chat messages generally.

i. Prof. Hochstetler opines that "the five individuals represented in the Log Dataset produced by Google did not personally switch" history on during the period covered by the Log Dataset "as [he] understand[s] they were instructed by Google to do."[15] This conclusion is meaningless because, as Prof. Hochstetler concedes elsewhere in his report, Chat messages will be preserved if history is on when a custodian joins a group (whether or not the custodian "personally" changes it).[16] Even assuming it is true for the sake of argument, Prof. Hochstetler's analysis confirms that none of the five custodians turned history to "off" when another Googler had switched it on.[17] Moreover, Prof. Hochstetler's report indicates that he did not in fact consider the litigation hold instructions and, moreover, he has misstated their content.

j. Prof. Hochstetler opines that a substantial percentage of chat conversations "had chat history turned off at least for some time during the timeframe covered by the Log Dataset."[18] Prof. Hochstetler's conclusion is meaningless

---

[12] Hochstetler Report at ¶ 8.
[13] GOOG-AT-MDL-C-000017436 at n.3.
[14] Hochstetler Report at ¶ 7(a).
[15] Hochstetler Report at ¶ 7(b).
[16] Hochstetler Report at ¶ 41 (Figure 4).
[17] Hochstetler Report at ¶ 40.
[18] Hochstetler Report at ¶ 7(c).

because it ignores the substantial evidence in the record regarding the ways in which Google employees use the Chat tool, and their varying roles and responsibilities. Considering Prof. Hochstetler's disregard of key evidence in the record, I consider his characterization of the Chat tool as being "used internally throughout Google for business-related communication" to be misleading or at best critically incomplete.[19]

k. Prof. Hochstetler opines that "relevant" messages were not retained because "chat conversations produced in this litigation had the chat history Retention Setting set to 'off' at some point[]" in the Log Period.[20] However, Prof. Hochstetler, who is not a lawyer, has defined relevance to include documents produced in a different litigation that were explicitly "unrelated to any party's claim or defense in this action."[21] By relying on chat logs—that were produced in a different litigation (with few overlapping custodians)—and ignoring other instances where chat history may have been appropriately turned on or off, Prof. Hochstetler has introduced a selection bias that renders his conclusion unreliable.

l. Prof. Hochstetler's analysis is also inconsistent with my empirical analysis of the documents produced in this case. Though Prof. Hochstetler provided the source code of scripts that he used for his analysis, he provided no detailed description of his methodology—for example, how his environment was set up, or how the CSV files containing the chat log data were organized on his computer. As a result, it is not possible to reproduce exactly the steps he took in undertaking his analysis. However, even assuming the truth of Prof. Hochstetler's analysis of the *Play* logs, the conclusions that he draws from that analysis are unreliable and undermined by the content and characteristics of the chat conversations produced in the case.

---

[19] Hochstetler Report at ¶ 9.
[20] Hochstetler Report at ¶ 7(d).
[21] *See* May 26, 2023 Letter from Robert J. McCallum to Kelly D. Garcia at 3-4.

HIGHLY CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

m. For all of the aforementioned reasons, and as further described below, all of Prof. Hochstetler's opinions referenced above are flawed and unreliable.

IV.   BACKGROUND

A.   Google's Chat Tool

11. Google's chat tool (hereafter, "Chat") is one of many methods used by Google employees (referred to within Google as "Googlers") to communicate. Googlers also routinely communicate via Gmail (an email tool), through face-to-face discussions, and by creating and sharing documents, including exchanging comments in those documents.[22]

12. Googlers use Chat for a variety of purposes. In addition to communicating about business-related topics (both substantive and non-substantive), Googlers use Chat to discuss a range of non-business topics, such as birth announcements, making lunch plans, and sending congratulatory or birthday wishes.[23]

B.   The *Play* Logs at Issue in the Hochstetler Report

13. The Play Store litigation was a separate lawsuit in the Northern District of California brought against Google regarding Google's Play store (*In re Google Play Store Antitrust Litigation*, No. 21-md-02981-JD (N.D. Cal.) ("*Play*")). The plaintiffs in the *Play* case alleged that Google monopolized the Android mobile app distribution market and did not make allegations concerning advertising, let alone Google's ad tech products. This is very different from the conduct at issue in this case.

---

[22] *See, for example*, Deposition of ███████, May 1, 2024 (hereafter "███████") at 307:16-308:3 ("Most of the communication is done either face-to-face or through documents, e-mails."); Deposition of ███████, April 12, 2024 (hereafter "███████") at 63:25-64:1 (agreeing that email, chats, and in-person discussions are interchangeable); *see generally* Deposition of ███████, GOOG-AT-MDL-007176468 - GOOG-AT-MDL-007176733, April 22, 2021 at 39:20-25 ("[C]hats [are] . . . part of our Gmail system".).

[23] *See* Deposition of Genaro Lopez, May 17, 2024 (hereafter "Lopez Dep.") at 105:23-106:25; 228-229; *See also, e.g.*, Deposition of ███████, April 17, 2024 at 185:3-12 ("I think occasional use, I think in general, yes, for kind of quick exchanges, like, are we going for lunch."); Deposition of ███████, November 14, 2023 at 229:9-14 ("I use chat for the situation that I described, that, okay, I'm late in the meeting, or where should we go for lunch, and those things."); Deposition of ███████, April 26, 2024 at 74:18-75:2 ("[Chat] tends to be a little bit more informal in terms of, you know, people sharing 'happy birthdays' or, you know, 'congratulations.'); Deposition of ███████, November 16, 2023 (hereafter "███████") at 243:7-11 ("We use chat to send, like, one-line messages. We don't use chat for material work discussions. It's usually like, Hey, are you coming to this meeting? Here is the link to the document. Things like that.").

HIGHLY CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

14. The "*Play* logs"[24] that are the subject of Prof. Hochstetler's report are a subset of logs produced in the *Play* case. The *Play* logs were not intended to measure the frequency with which any users toggled "history on" or "history off". Rather, the *Play* logs were used for internal debugging purposes.[25] The fields included in the debugging log include the time a message was sent, the unique user ID of the Google employee sending the message, the "space_id" that identifies a Chat conversation or thread, and "actions," one of which corresponds to a user changing their history setting.[26] But the *Play* logs do not show the contents of any Chat messages.[27] Prof. Hochstetler concedes that the *Play* logs are not exhaustive and do not appear to capture all Chat messages.[28]

C.   Google's Collection and Production of Electronically Stored Information

15. In my experience, large modern organizations (such as Google) typically store data in a variety of different formats. In this litigation, Google has produced over 6.3 million documents, over 750 terabytes of data, and over 80 gigabytes of source code.[29] I understand that the more than 6.3 million "documents" comprises emails, Chat messages, spreadsheets, presentations, hard copy documents, memoranda, notes, dashboards, comment notification e-mails, and text messages.

16. I understand that Google produced chat messages by exporting them from Google Vault, an information governance and discovery tool. Google Vault exports

---

[24]   Prof. Hochstetler refers to the *Play* logs of the five custodians as the "Log Dataset." Hochstetler Report at n.1 and ¶ 7. I use both terms interchangeably.

[25]   Oct. 1, 2024 Email from Veronica Bosco to Ethan Glenn.

[26]   Sep. 23, 2024 Email from Veronica Bosco to Ethan Glenn. Prof. Hochstetler's contention that "if a space_id field is missing in the Log Dataset, it is unrecoverable," is inaccurate. Hochstetler Report at ¶ 30. Google informed Plaintiffs that "there are several reasons why a 'space_id' might be missing for a logged action", including, "[f]or example, 'space_id' is not logged for actions that are specific to a user, but aren't associated with a specific space." Sep. 23, 2024 Email from Veronica Bosco to Ethan Glenn. Further, because the log is a debugging log, "a 'space_id' might not have been logged if it was not necessary for debugging purposes", meaning that certain space_id data was never logged in the first instance and "it does not make sense that the data would be available elsewhere." Oct. 1, 2024 Email from Veronica Bosco to Ethan Glenn. Accordingly, Prof. Hochstetler's contention that Google deleted data when that data were not logged in the first place is misleading.

[27]   Aug 29, 2024 Email from Robert J. McCallum to Ethan Glenn; *see also* Hochstetler Report at ¶ 23.

[28]   Hochstetler Report at ¶ 27 ("the logs entries do not appear to be exhaustive in received messages" and "contain a number of actions that are not related to sending and receiving messages, for instance, muting a conversation, marking a message as read, or filtering messages.").

[29]   *See* Google's Opposition to Plaintiffs' Motion to Compel Written Discovery Regarding Google Chats, *The State of Texas, et al. v. Google, LLC*, The United States District Court, Eastern District of Texas, Sherman Division, Case No.: 4:20-cv-957-SDJ, July 5, 2024 (hereafter "ECF No. 545") at 4-5.

HIGHLY CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

communications from Google Chat as "conversations" which may contain multiple individual messages between and among Chat participants in a single document.[30] The sending or receipt of an individual chat message does not necessarily lead to the export or production of a separate document. A single exported Chat conversation can contain hundreds or thousands of individual chat messages. Additionally, when no messages are sent within a 24 hour period, a Chat conversation is created and saved as a single document.

17. I understand from counsel that these documents were collected from custodians identified by Google as potentially having documents responsive to Plaintiffs' requests for production. Google conducted searches of search terms agreed upon with Plaintiffs. Google's counsel then reviewed for responsiveness and produced responsive documents to Plaintiffs.

18. Plaintiffs have received the benefit of 202 custodians agreed across this case, as part of the multi-district litigation with which this litigation was consolidated before being remanded back to Texas, and as part of a separate litigation brought by the Department of Justice in the Eastern District of Virginia (the "Virginia Case").[31]

19. I understand from counsel that in the Virginia Case, Google also agreed to run search terms that focus on Googler's use of the Chat history settings, but which are otherwise not relevant to the claims and defenses in the case (the "Discovery-on-Discovery Search Terms").[32] Those Discovery-on-Discovery Search Terms are set forth in Schedule 3. I understand from counsel that Google produced the resulting chat messages in both the Virginia Case and this case, regardless of whether they were relevant to the claims and defenses.

20. Prof. Hochstetler states that "by the beginning of 2022, 141 Google employees had been placed on litigation hold."[33] This statement is not accurate as written. I understand that Prof. Hochstetler cites a letter in which Google provided information about legal holds placed for Google employees *who were agreed-upon custodians,* many of whom

---

[30]   *See, generally,* "Vault export contents," Google Vault Help, https://support.google.com/vault/answer/6099459?sjid=11870055449658323261-NA#zippy=%2Cexport-contents.

[31]   *United States, et al. v. Google LLC*, No. 23-cv-00108-LMB-JFA (E.D.V.A.).

[32]   *See, for example,* May 26, 2023 Letter from Robert J. McCallum to Kelly D. Garcia at 3-4.

[33]   Hochstetler Report at ¶ 5 and n.4.

HIGHLY CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

were placed on hold after Prof. Hochstetler's arbitrary cut-off at the beginning of 2022 and therefore are not covered by his analysis.[34] Prof. Hochstetler also ignores that Google placed many employees on hold (and thus suspended normal preservation processes for their documents), even if they were not agreed-upon custodians in this case.[35]

D.   The Retention Periods at Issue in This Case

21. I understand that Google adopts different retention periods for different types of communication based on the substantive business value of the communication. Google's 30(b)(6) representative Mr. █████████ testified that Google accounted for the "substantive business value" of "different types of communications" in establishing "different policy treatment" for different means of communication.[36] Prof. Hochstetler suggests throughout his report that Google's information governance practices are unusual or cannot be explained by Google's business judgment alone.[37] But, as I explain below, these suggestions lack basis because these kinds of information governance practices are widespread in my experience. Indeed, I understand that the record in this case reflects that many Plaintiff States have adopted similar information governance practices.

22. For example, many of the Plaintiff States have adopted retention policies grounded in similar principles that distinguish retention periods based on the value of the information at issue. A few of these are listed below.

- Texas's Records Retention Schedule distinguishes between "administrative" correspondence (which is retained for 4 years), "general" correspondence (which is only retained for 2 years) and "no action required" ("require no action to be taken" and has a "purge date set by internal division policy.").[38]

---

[34]   Hochstetler Report at n.4.

[35]   *See* Hearing Transcript, *The State of Texas, et al. v. Google, LLC*, The United States District Court, Eastern District of Texas, Sherman Division, Case No.: 4:20-cv-957-SDJ, July 10, 2024 (hereafter "Hearing Transcript") at 57:18-21 ("There's a distinction between collection and preservation . . . These are just the ones that were collected. There would be more that would be preserved.").

[36]   Lopez Dep. at 123:18-124:1.

[37]   *See, for example,* Hochstetler Report at ¶ 13 ("Google Chat users are able to turn history 'on' or 'off' for a conversation at any time unless this capability is restricted by the enterprise's administrator.").

[38]   TXOAG-0009754 at 65-66.

HIGHLY CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

- The document retention schedule for Utah's Division of Consumer Protection distinguishes between "housekeeping correspondence" (which is only kept for 2 years) and "minutes" (which are kept for 1 year).[39]

- The document retention schedule for Indiana's Unclaimed Property Division—the Division that purchases advertising within the Indiana OAG—has a ten year retention policy, but which only applies to three categories of documents: closed claim files, annual reports of abandoned property, and annual statements that no property has been abandoned.[40]

23. In fact, it is normal and sensible for companies to have data management policies that permit routine deletion of data. For example, the Sedona Conference, a nonprofit organization focused on advancing legal and policy research, emphasizes that organizations should balance the business value against the cost and risk associated with retaining information in Principle 6 of the Sedona Conference Commentary on Information Governance.[41] The Sedona Conference further states that "[e]phemeral messaging can assist with implementation of the life-cycle process by eliminating data with no ongoing business value, particularly since a sizeable portion of the data growth involves this type of information (e.g., routine communications, meeting requests, duplicative email chains to large groups, etc.)."[42]

24. Prior to February 8, 2023, the ordinary course retention period for Google Chat depended on the history setting of a particular message, but in all instances, all "history on" Chat messages were retained for at least 30 days (and many were saved for 18 months).[43] During this period, any "history on" messages subject to legal hold were retained for the duration of the hold.

25. On February 8, 2023, the settings for all Google Chat messages were automatically defaulted to "history on" and the ordinary course retention period varied from 30 days

---

[39] *See, for example,* Utah Department of Government Operations, Division of Archives and Records Service: Department of Commerce; Division of Consumer Protection, https://tinyurl.com/mt6hh8ue (last accessed Nov. 25, 2024).

[40] *See* INOAG-GOOG_000000040.

[41] *The Sedona Conference Commentary on Information Governance, Second Edition*, 2019 at 139-141.

[42] *The Sedona Conference Commentary on Ephemeral Messaging*, 2021 at 455.

[43] *See* GOOG-AT-MDL-009709510 (Gmail Retention Policy Aug. 21, 2015–Feb. 25, 2021); GOOG-AT-MDL-009709520 (Google Chat Retention Policy Nov. 18, 2020–Feb. 25, 2021); GOOG-AT-MDL-009709508 (Google Chat Retention Policy Feb. 26, 2021–Sep. 30, 2021); GOOG-AT-MDL-009709522 (Google Chat Retention Policy Oct. 21, 2021–Dec. 7, 2021); GOOG-AT-MDL-009709506 (Google Chat Retention Policy Feb. 26, 2022–May 16, 2023); GOOG-AT-MDL-009709518 (Google Chat Retention Policy May 17, 2023–Present).

HIGHLY CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

to 18 months, depending on the type of message. Also, for any Chat conversation involving an individual who was subject to a litigation hold, such setting could not be toggled to "history off" and those messages were retained for the duration of the hold.[44] Google's corporate representative has testified in this case that this "Google-specific change" required "hundreds of hours to develop and deploy" across "multiple teams."[45]

26. Prof. Hochstetler notes that "Google's systems did not automatically enforce preservation of chat conversations."[46] I understand that many of the Plaintiff States—including Florida, Arkansas, Kentucky, and North Dakota—have admitted that they similarly do not automatically enforce the preservation of all chat conversations.[47]

27. In his report, Prof. Hochstetler states, "[i]nterestingly, prior to [November 2020], the chat retention policy for a new Google Chat room (as opposed to a chat room started in Google Hangouts), was indefinite."[48] Prof. Hochstetler does not identify why that change is "interesting" to him, but the implication of the statement is that the shortening of the retention period was nefarious in some unspecified way.

28. At least one Plaintiff State made changes to its chat retention policy in the same period. On September 16, 2020, North Dakota Information Technology announced that "a new Microsoft Teams Chat deletion policy will be implemented," and that, "[o]nce implemented, any chats with a date prior to September 18, 2019, will be deleted."[49] The change in deletion policy extended to all "Teams Chats (including one-on-one chats, meeting chats, and group chats)."[50] The COVID-19 pandemic forced many

---

[44] By contrast, some Plaintiff States, like South Carolina, permanently delete instant messages after ten days. *See* South Carolina Office of Attorney General Office Manual, Section 4 at 9.

[45] Lopez Dep. at 138:6-15. In light of Mr. Lopez's testimony, Prof. Hochstetler's assertion that "it would not have been difficult for Google to restrict its own employees using Google Chat from changing Retention Settings at an automated level" is misleading. Hochstetler Report at ¶ 18. While Google's Chat tool permits administrators to change the history setting for all users within an organization, Mr. Lopez testified that "it's much more difficult to only have that change apply to a subset of all the employee population." Lopez Dep. at 156:17-22.

[46] Hochstetler Report at ¶ 7(b).

[47] *See* Sep. 18, 2024 Letter from Jonathan P. Wilkerson to Robert J. McCallum at 4 (the OAG of the State of Arkansas "does not . . . have a document retention policy," and the OAG for the State of Florida does not specifically preserve chat conversations, as the OAG preserves documents "based on the content and purpose of the record, not the format or technology used"), 7 ("Concerning other correspondence unrelated to official business, the generally applicable Kentucky State Government Records Retention Schedule provides that such 'nonbusiness related correspondence' should not be retained."), 8 ("North Dakota did not suspend the auto-deletion of Teams Chat messages because it was determined not necessary to preserve any potentially relevant documents.").

[48] Hochstetler Report at ¶ 14.

[49] NDAG-ADTECH-000008805 at -806.

[50] NDAG-ADTECH-000008805 at -806.

HIGHLY CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

employees in the U.S. to work from home, resulting in an increased reliance on digital communication tools, such as chatrooms and other messaging tools, to maintain efficient collaboration.[51] As a result, changes in data policies of this nature were not uncommon in this timeframe.[52]

29. Overall, Prof. Hochstetler's analysis of Google's retention settings is flawed because he describes those retention settings incorrectly. He states, "Retention Settings are applied per conversation."[53] This statement is wrong. Retention settings are applied on a *per message* basis.[54] As a result, Prof. Hochstetler's description of this feature is mistaken through the rest of his analysis.[55]

30. Additionally, Prof. Hochstetler omits an important detail regarding retention rules. While he states that "retention rules within Vault were to be suspended if a litigation hold was issued,"[56] he neglects to specify that such a suspension only applies to history on messages. There is no technical method to preserve messages sent in a "history off" state.[57] I discuss various flaws in Prof. Hochstetler's analysis in Section VII below.

---

[51]   Sigal Kordova and Ron S. Hirschprung, "Effectiveness of the forced usage of alternative digital platforms during the COVID-19 pandemic in project communication management," *Heliyon 9*, no. 11 (2023), at 1-2, 12.

[52]   *See, for example,* Shauna Sowersby, "Why are Washington state agencies destroying Microsoft Teams chat messages," *The Olympian*, Updated October 13, 2023, https://www.theolympian.com/news/politics-government/article280170064.html; Amazon Web Services, "AWS announces custom chat retention policies for the Amazon Chime application," May 21, 2020, https://aws.amazon.com/about-aws/whats-new/2020/05/aws-announces-custom-chat-retention-policies-for-the-amazon-chime-application/. See also "Microsoft Teams personal chats removed after 90 days," The University of British Columbia, March 31, 2021, https://lthub.ubc.ca/2021/03/31/microsoft-teams-90-day-retention/.

[53]   Hochstetler Report at ¶ 15.

[54]   *See* ▇▇▇ Dep., Ex. 416 at 52:17-20 ("[W]hen history on is turned on, it applies only to *messages* sent after that setting change.") (emphasis added).

[55]   Hochstetler Report at ¶ 21 ("[D]isabled changes to the Retention Setting for conversations"); ¶ 40 ("[N]one of the five individuals represented in the Log Dataset personally changed the chat history Retention Setting of any of their conversations during the Log Period."); ¶ 410 ("[T]he Retention Setting is applied per conversation."); Section VI ("Almost 90% of conversations (Groups) had retention settings toggled off at some point."); ¶ 66 (conflating conversations and 'Groups' and implying that Retention Settings are applied to conversations); ¶ 81(a) ("The proportion of these chat conversations for which the Retention Setting was 'off' for the entire portion").

[56]   *See* ▇▇▇ Dep. at 78:10-19 ("for a chat to be retained at all by Google [Vault]", the chat "has to be sent in a history on state"); 73:17-25 ("Q. And so if a litigation hold has not been issued as to an individual, there will be no retention of that individual's chats? A. They'll continue to be retained in the normal course of business via our regular rules, but there will not be any special treatment as you would see in the legal hold.").

[57]   *See* ▇▇▇ Dep. at 77:7-16 (confirming that "history-off messages are never available to Vault . . . they will disappear from the user's view after 24 hours, but they are never available for retention purposes.").

HIGHLY CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

E.   **The Litigation Holds at Issue in This Case**

31. Beginning in September 2019, Google issued litigation hold notices to numerous Googlers in connection with Plaintiffs' investigation and subsequent litigation. Those litigation hold notices instructed Googlers to "not use" Google Chat or any other "messaging apps to discuss any topics covered" by their legal holds, but that, if they "must do so," to "make sure the settings preserve the messages, such as switching 'history on.'"[58] As described further below at Section VI.B.iii, Prof. Hochstetler did not list these litigation hold notices in his materials relied upon or considered.

V.   SUMMARY OF THE OPINIONS AND ANALYSIS PERFORMED BY PROF. HOCHSTETLER

A.   **Summary of Prof. Hochstetler's Analysis of the Log Dataset**

32. Prof. Hochstetler states that he has been "asked by counsel for the State of Texas, on behalf of all Plaintiff States in this case, to analyze chat log data" to 1) determine how many messages were sent by certain Google employees daily, 2) quantify the degree to which Chat messages were not retained in these conversations and the frequency with which the chat history retention setting ("Retention Setting") was switched on or off, and lastly 3) examine the chat logs in conjunction with the Chat messages produced in this case to assess whether the provided conversations were impacted by the absence of Chat message retention.[59] To assess these three topics, Prof. Hochstetler analyzed a dataset containing logs of Google Chat messages for five individuals ("Log Dataset") over a period of 68 days ("Log Period").[60] I understand that this Log Dataset that Prof. Hochstetler analyzed contained 307,473 Google Chat log entries spanning December 9, 2022 through February 14, 2023.[61]

33. Prof. Hochstetler starts his analysis by determining the types of actions for which the <retention_state> field is populated and are related to sending and receiving messages.[62] He finds eight distinct actions for which the <retention_state> field is

---

[58]   *See, for example*, GOOG-AT-MDL-C-000087449; GOOG-AT-MDL-C-000087452; GOOG-AT-MDL-C-000087456.
[59]   Hochstetler Report at ¶ 2.
[60]   Hochstetler Report at ¶ 7.
[61]   Hochstetler Report at ¶ 25.
[62]   Hochstetler Report at ¶ 37. I understand that the <retention_state> field indicates whether the Retention Setting was toggled on or off. Hochstetler Report at ¶ 35.

HIGHLY CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

specified and that are "[a]ssociated with sending and receiving messages."[63] Using the Google documents produced in this case, he identifies that out of these eight actions, two "key actions" he claims are associated with the status of the Retention Setting, namely <.DynamiteFrontendService–UpdateGroupRetentionSettings>, which updates the retention settings of a group chat, and <DynamiteService–UpdateGroupNotificationSettings>, which updates the notification settings for a user on a group chat.[64] Prof. Hochstetler uses the former action to check "whether any of the 5 individuals represented in the Log Dataset changed the retention settings for a Group chat," and if so, to identify the person who modified the setting and how often the changes occurred.[65] As I note in the subsequent section, he finds that none of these five Googlers changed the Retention Setting themselves during the Log Period.[66]

34. Next, Prof. Hochstetler attempts to determine the volume of messages sent by the five Googlers during the Log Period. To estimate what supposedly represents the total number of messages sent per person, he counts the number of backend logs, which are the actions starting with <DynamiteService>.[67] He then calculates the average number of messages sent per day by dividing each person's total message count by the number of days when the person sent at least one message.[68]

35. Prof. Hochstetler then attempts to quantify the actions with the history retention enabled for each of the five individuals separately, as well as collectively.[69] He does this by first filtering the Log Dataset to actions that included a <retention_state> key in the <dynamite_server_extension>[70] dictionary. For each action group, he counts how many actions have <retention_state> set to "EPHEMERAL_ONE_DAY" versus

---

[63]   Hochstetler Report at ¶ 37 (Table 5).

[64]   Hochstetler Report at ¶ 39 (Table 6). Regarding the latter action identified in Google's documentation, <DynamiteService–UpdateGroupNotificationSettings>, Prof. Hochstetler concludes that "notifications setting changes do not have an obvious connection to retention changes" and therefore, opts to not use this field in his analyses. Hochstetler Report at ¶ 45. While he did not find any chat logs corresponding to these actions, Prof. Hochstetler noted that he also identified two additional actions that appeared to be related to chat retention. *See* Hochstetler Report at ¶ 42.

[65]   Hochstetler Report at ¶ 40.

[66]   Hochstetler Report at ¶ 40.

[67]   Hochstetler Report at ¶ 46.

[68]   Hochstetler Report at ¶ 46 (Table 7).

[69]   Hochstetler Report at Section V.E.

[70]   Prof. Hochstetler describes the <dynamite_server_extension> field as "[c]ommon metadata from frontend and backend server log messages of Google Hangouts." *See* Hochstetler Report at ¶ 28 (Table 3).

HIGHLY CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

"PERMANENT."[71] Relatedly, "EPHEMERAL_ONE_DAY" refers to "history off" and "PERMANENT" refers to "history on."[72]

36. In his Table 9, he reports, for each of the five individuals, how many messages sent were not retained during the full Log Period and during the timeframe before February 8, 2023, "when Google changed the default Retention Setting to PERMANENT."[73] However, he notes that because there are typically three log entries for each received message, the actual number of received messages "is likely close to one third of the number of logs reported" in the last two columns of his Table 9.[74]

37. Using a similar methodology, Prof. Hochstetler also attempts to assess the status of the <retention_state> field in the unique Groups in which the five Googlers participate. For each of these five individuals, he counts the number of unique Groups 1) that the person is a part of in the Log Dataset, 2) where the <retention_state> was "off" during the Log Period, and 3) where the <retention_state> was changed from "off" to "on" (or *vice versa*) at some point during the Log Period.[75]

B.   Prof. Hochstetler's Opinions

38. Overall, Prof. Hochstetler opines that the total number of messages that were not retained by the 141 Googlers who were subject to a litigation hold in 2022 was "a million and a half or more."[76] Regarding the five Googlers in the Log Dataset, Prof. Hochstetler concludes that "the vast majority of messages sent and received" (over 87 percent) by these five individuals were sent and received while the history retention setting was off.[77] Specifically, Prof. Hochstetler claims that 9,592 messages that were sent and "at least" 8,974 messages that were received by these five employees in the Log Dataset were lost during the Log Period.[78] However, he finds based on his analysis of the Log Dataset that these five employees did not personally change the Retention

---

[71]   Hochstetler Report at ¶ 47 (Table 8).
[72]   Hochstetler Report at ¶ 29 (Table 4).
[73]   Hochstetler Report at ¶ 48 (Table 9).
[74]   Hochstetler Report at n.92.
[75]   Hochstetler Report at Section VI.
[76]   Hochstetler Report at ¶¶ 8, 82.
[77]   Hochstetler Report at ¶ 36.
[78]   Hochstetler Report at ¶ 81(b).

HIGHLY CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

Setting of any of their conversations during the Log Period and that any changes to Retention Settings in their conversations were carried out by other chat participants.[79]

VI.   EVALUATION OF PROF. HOCHSTETLER'S METHODOLOGY, ANALYSIS, AND OPINIONS

A.   Prof. Hochstetler's Analysis Is Not Statistically Significant

i.   The Importance of Statistical Significance

39. Statistical significance can help assess the reliability of conclusions drawn from data. Statistical significance is used to test whether results based on a set of data are unlikely to happen merely due to chance.[80] Reporting statistical significance helps ensure the conclusions drawn from data are accurate and trustworthy.

40. Prof. Hochstetler does not assert, in his report, that the conclusions that he draws from his limited sample are statistically significant. As further described below, the 5-person sample used by Prof. Hochstetler is biased and small, and thus not representative of all the agreed upon custodians subject to a litigation hold at the beginning of 2022.[81] Therefore, the conclusions that Prof. Hochstetler draws from it are not reliable.

ii.   The Play Logs Represent an Unreliable Sample of Convenience

41. The *Play* logs produced in this case reflect a subset of logs produced in the *Play* case. No recognized statistical methodology was used to identify the *Play* Logs and no methodology was used to ensure that the sample was representative or non-biased (or to assess potential bias in the sample). Instead, Plaintiffs sought production of the *Play* logs largely on the basis that they had already been produced in the *Play* case and so there was no burden to Google in reproducing them in this case.[82] I understand that

---

[79]   Hochstetler Report at ¶¶ 40-41.

[80]   *See* David H. Kaye and David A. Freedman, "Reference Guide on Statistics," in REFERENCE MANUAL ON SCIENTIFIC EVIDENCE, 284-285 (3rd ed. 2011), (hereafter "2011 Kaye & Freeman") at p. 251; David S. Moore et al., INTRODUCTION TO THE PRACTICE OF STATISTICS (6th ed., 2007), at p. 184; Amy Gallo, "A Refresher on Statistical Significance," *Harvard Business Review*, 2016, https://hbr.org/2016/02/a-refresher-on-statistical-significance.

[81]   Hochstetler Report at n.4 and n.106. In his report, Prof. Hochstetler extrapolates his Opinions to the 141 employees subject to a litigation hold, writing that "[m]y analysis leads me to conclude that it is reasonable to assume that the number of messages lost per year, across all employees under the litigation hold, was close to 1.5 million." *See* Hochstetler Report at ¶ 82.

[82]   *See* Hearing Transcript at 84:24-85:2.

Google's counsel resisted production of the *Play* logs on the basis that, among other things, there is "very, very little overlap between the custodians in the Play case and in this case," and that the *Play* logs would only "cover a sliver of time in early 2023."[83] As a result, Google argued, any attempt by Plaintiffs "to approximate how frequently custodians in the *Play* case switch their history on or off" would be "meaningless" in this case.[84]

42. The data reflected in the *Play* logs represents what is known as a "sample of convenience". A sample of convenience references a sample that is drawn from readily accessible subgroups within a larger population.[85] Although samples of convenience are relatively easy to gather, they may in some circumstances suffer from serious biases if special precautions are not employed when designing the sample selection methodology. As set forth below, the sample of convenience used by Prof. Hochstetler here is vulnerable to selection bias and, as a result, is underinclusive and thus unrepresentative of the larger population of interest.[86]

43. Prof. Hochstetler relies on a mere five custodians over 68 days to estimate the number of messages allegedly lost by more than 140 custodians per year.[87] Prof. Hochstetler also adjusts his calculation on the unsupported assumption that "half of the employees under litigation hold had left the company by 2022."[88] It is improper to extrapolate from such a small and unrepresentative sample to all employees subject to a litigation hold. First, as I discuss in Section VI.B, the five custodians examined may use Chat in ways that systematically differ from the Chat usage of all employees subject to a litigation hold. Second, as I discuss in this section and Section VI.B.iv, the five custodians (on which Prof. Hochstetler focuses) are not representative of the larger population of employees (who are subject to a litigation hold) as to the relevance of

---

83   *See* Hearing Transcript at 75:18-76:1.
84   *See* Hearing Transcript at 75:3-17.
85   Shari Seidman Diamond, "Reference Guide on Survey Research," in REFERENCE MANUAL ON SCIENTIFIC EVIDENCE, 3rd ed. (Washington, DC: The National Academies Press, 2011), at p. 419; 2011 Kaye & Freedman at p. 285; Paul J. Lavrakas and Jeffery A. Stec, "Survey Research in Litigation" in LITIGATION SERVICES HANDBOOK: THE ROLE OF THE FINANCIAL EXPERT (eds. Roman L. Weil, Daniel G. Lentz, & Elizabeth A. Evans, 6th ed. 2017), at p. 18.
86   2011 Kaye & Freedman at pp. 224-225; TRADEMARK AND DECEPTIVE ADVERTISING SURVEYS, 2nd ed., Eds. Shari Seidman Diamond and Jerre B. Swann (American Bar Association, 2022), at p. 70.
87   Hochstetler Report at ¶ 82 and n.106.
88   Hochstetler Report at n.4.

HIGHLY CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

this particular lawsuit to their roles at Google. Third, given the facts of this case, the Log Period is likely to be too short—68 days does not yield a sample large enough from which to draw conclusions applicable across multiple years. Fourth, as Prof. Hochstetler concedes, the Log Period covers a popular holiday period during which it is plausible that Chat usage patterns may differ from the norm.[89]

iii.   *The Conclusions Drawn by Prof. Hochstetler Are Unreliable*

44. Prof. Hochstetler opines that there were an "estimated 18,566 messages lost" for five custodians over a 68 day period.[90] This sample, on which Prof. Hochstetler relies, is a biased sample of convenience that is not representative of the larger population of employees (who are subject to a litigation hold). Thus, it is my opinion that the sample used by Prof. Hochstetler is not a proper basis to extrapolate and the result of any such extrapolation is unreliable. By Prof. Hochstetler's own calculation, his reliance on this improper sample results in "1.4 million messages lost in a single year."[91] But Prof. Hochstetler goes one step further and adds an arbitrary and unexplained 100,000 messages to that figure.[92] It follows that Prof. Hochstetler's opinion—"that the total number of messages not retained by Google employees subject to a litigation hold was a million and a half or more in 2022 only"[93]—is unreliable.

B.   **Prof. Hochstetler's Analysis Ignores the Discovery Record in This Case**

i.   *The Importance of Using the Data and Information Available*

45. In my experience, if there is available a superior or otherwise more reliable source of data (e.g., sworn testimony or independent research studies), that data source should be used over–or at least in conjunction with and as a check on–less reliable alternatives to inform the assumptions concerning one's analyses. Using more reliable forms of data reduces the risk of errors, and ensures that one's conclusions are more accurate and credible.

---

[89]   Hochstetler Report at ¶ 8 and n.4.
[90]   Hochstetler Report at n.4.
[91]   Hochstetler Report at n.4.
[92]   Hochstetler Report at ¶ 8.
[93]   Hochstetler Report at ¶ 8.

ii.   *Prof. Hochstetler's Analysis Does Not Account for the Way Googlers Use Chat*

46. Prof. Hochstetler's report does not consider the substantial amount of information in the discovery record of this case addressing the manner in which individual Googlers do (and do not) use Google's Chat tool.

47. Aside from discussing substantive business, Googlers also use Chat for many other purposes. I have categorized these non-substantive other uses according to testimony in the record as work-related logistics, other non-substantive work-related communication, and non-work-related communications. I provide examples of each category below.[94]

a.   Work-related logistics, such as organizing meetings

48. Several Googlers testified that they used Google's Chat tool primarily for work-related logistics (e.g., organizing meetings and following up on tickets).

49. Scott Sheffer, a Vice President of Global Monetization at Google, testified that "when I communicate over chat, it's mostly, hey, can you jump into this meeting, or are you going to make it on time to the next meeting, or something like that."[95]

50. ██████████████████████████████████, testified that he "mainly used Chats to set up meetings."[96] Another Google employee, ██████████, who is ████████████████████████, testified that he used Chat for communicating "[l]ogistic-type of information" such as discussing meeting topics.[97]

51. Additionally, Neal Mohan, the Chief Executive of YouTube and a Senior Vice President at Google, testified he generally uses Chat for scheduling meetings or checking in on another employee.[98] Another Google employee, ████████████████ who is ████████████████████████, testified that he and his coworkers "don't use

---

[94] The titles indicated for each Googler reflect their titles at the time of their depositions or their cited document.

[95] Deposition of Scott Sheffer, GOOG-AT-MDL-007177648 - GOOG-AT-MDL-007178018, July 20, 2021, at 23:24-24:13, 35:13-22.

[96] Deposition of ██████, April 1, 2024 at 27:21-28:4 and 35:6-36:8 ("Q. Okay. So you mainly just used Chats to set up meetings to discuss? A. Yes.").

[97] Deposition of ████████, April 30, 2024 at 78:5-7 and 309:17-20 ("An example would be, reminder, we have a team meeting tomorrow, does anyone have any topics. Logistic-type of information.").

[98] Deposition of Neal Mohan, October 30, 2023 at 37:20-24 and 108:12-17 ("I occasionally use Chats. Oftentimes to schedule meetings or check in on an employee, that kind of thing.").

HIGHLY CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

chat for material work discussions" and that Chat is usually used for logistics like meeting organization, following up on tickets, and sometimes HR-related questions.[99]

b. Other non-substantive work communication

52. Employees also use Chat for other non-substantive work communication.

53. For instance, a Google document shows Google employee ████████████ initiated a chat conversation by saying "Team chat! All fun stuff, gossip, pictures you wanna share! LETS BUILD SOME OF THAT TEAM SPIRIT!" [100] Another Google employee, ████████████, ████████████████████████████, testified that, among other things, he used Chat to provide a colleague a "quick answer."[101]

c. Non-work-related communication

54. Moreover, employees also use Chat for non-work-related conversations.

55. According to the testimony of ████████, ████████████████████, most work related communication on AdX was "done either face-to-face or through documents, e-mails."[102] ██████ testified that "[m]ost of the chat conversation was on the short side, like, let's go to lunch or something."[103]

56. Further, Aparna Pappu, a Vice President and General Manager of Google Workspace, described some chats as "ad hoc" and "informal, kind of, not super relevant in some sense"; she further stated that it was a "force of habit, usually, to have substantive, kind of, conversations over e-mail and Chat was more chitchat" during her testimony.[104]

57. In addition, Prof. Hochstetler assumes that "Spaces," which he describes as "conversations [within Google Chat] focused on a specific theme, e.g., a project or a common area of interest," are used to discuss exclusively work themes.[105] He does not

---

[99]   ████████████ at 10:7-13, 242:9-252:5.
[100]   GOOG-AT-MDL-B-003980132 at -132.
[101]   Deposition of ████████████, May 15, 2024 at 35:13-18, 75:18-75:22 (using Chat to provide a colleague a "quick answer.").
[102]   ████████ at 22:21-23:10, 307:16-308:13.
[103]   ██████. at 307:16-308:13.
[104]   Deposition of Aparna Pappu, August 11, 2023 (hereafter "Pappu Dep.") at 30:18-31:8, 37:4-19, 289:14-20.
[105]   Hochstetler Report at ¶ 11.

HIGHLY CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

account for the fact that not only are not all Spaces work-related, but that many employees do not even use Spaces.[106]

58. In light of this variety of usage patterns, I consider Prof. Hochstetler's characterization of Google Chat as a tool that is "used internally throughout Google for business-related communication" to be incorrect and misleading.[107]

iii.   Prof. Hochstetler's Analysis Does Not Accurately Reflect the Litigation Holds

59. Many of Prof. Hochstetler's opinions are premised on the assumption that documents should have been retained because Googlers were instructed to retain documents pursuant to a "litigation hold."[108] However, as I explain below, Prof. Hochstetler's analysis omits the fact that Googlers' litigation holds instructed them not to use Chat to discuss topics covered by the litigation hold. This omission compounds the flaws I noted above in Section VI.B.ii: not only do many Googlers use Chat to discuss non-business-related topics even in the ordinary course of their workdays, but Googlers were also specifically instructed not to use Chat to discuss topics relevant to this litigation.

60. Prof. Hochstetler's understanding of a "litigation hold" appears to be drawn from a publicly available glossary defining the term.[109] Google produced redacted versions of the litigation holds issued in this case to Plaintiffs.[110] Based on his report,[111] it does not appear that Prof. Hochstetler considered or relied upon these documents.

61. Having not considered or relied on the litigation holds issued in this case, it is unsurprising that Prof. Hochstetler misstates their content. Prof. Hochstetler states that

---

[106]   *See, for example,* ▮▮▮▮▮▮ at 251:7-9 ("Q. Do you ever use Google Spaces at work? A. No."); Deposition of Kenneth Marco Hardie, November 14, 2023 at 120:14-18 (when questioned whether he uses "Google Chat or Google Teams or Spaces or any of the Google collaborative chatting applications," he answered only that he uses Chat); *see also* Deposition of ▮▮▮▮▮▮, May 1, 2024 at 323:5-9 ("A . . . . We have a black chat group, which is black leaders of Google. We have a St. John's group, which is the space St. John's terminal and all the people who actually sit within our area).

[107]   Hochstetler Report at ¶ 9.

[108]   *See, for example*, Hochstetler Report at ¶¶ 7(b)(c), 18, 21.

[109]   Hochstetler Report at n.32 (citing Thomson Reuters, "Glossary: Litigation Hold," https://uk.practicallaw.thomsonreuters.com/9-501- 9293?transitionType=Default&contextData=(sc.Default). Accessed on September 25, 2024.).

[110]   GOOG-AT-MDL-C-000087449; GOOG-AT-MDL-C-000087452; GOOG-AT-MDL-C-000087456; GOOG-AT-MDL-C-000087457.

[111]   Hochstetler Report at Appendices A and B.

HIGHLY CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

his "analysis shows that the five individuals represented in the Log Dataset produced by Google did not personally switch the setting to turn history 'on' for any (0%) of the conversations they participated in during the Log Period, per Google's documentation, *as I understand they were instructed by Google to do if they were subject to a litigation hold and discussed a relevant topic*."[112]

62. This misstates the instructions given to employees. In fact, employees were instructed to "not use" Google Chat or any other "messaging apps to discuss any topics covered" by their legal holds, but that, if they "must do so," to "make sure the settings preserve the messages, such as switching to 'history on.'"[113] I have reviewed deposition testimony in this case by Googlers discussing their usage of Chat. Numerous deponents showed evidence that they had an understanding of, and compliance with, the instructions given in the litigation holds as they pertain to Chat.[114] Further, no employee testified in their depositions that they ever deliberately deleted relevant communications, including Chat messages. For example, ███████, a Google employee who testified as a corporate representative and whose title is Product Go To Market, testified that she had never deleted any Chat messages and never been instructed to do so.[115]

63. Prof. Hochstetler also states that "[t]he recipients' compliance with the litigation hold was not monitored by Google."[116] Google informed Texas over four years ago in February of 2020 that Google sends legal hold recipients regular reminders of their

---

[112] Hochstetler Report at ¶ 7(b) (emphasis added); *see also* Hochstetler Report at ¶ 18 ("Lopez states that recipients of a litigation hold *were instructed that their chat messages needed to be preserved*, however, Google did not enforce automatic preservation of these messages until February 2023. Instead, it was left up to the recipients of the hold to identify relevant conversations and turn chat history 'on' during those relevant conversations.") (emphasis added).

[113] *See* GOOG-AT-MDL-C-000087449 at -451.

[114] *See, for example*, Deposition of Samuel R. Cox, GOOG-AT-MDL-C-000071500 - GOOG-AT-MDL-C-00007196, June 25, 2024 at 370:7-371:14 ("A. I believe when I was told that I was on legal hold that I turned my history on to make sure that it was on respecting the legal hold."); Deposition of ███████, GOOG-AT-MDL-007172452 - GOOG-AT-MDL-00717273, October 19, 2021 at 173:13-20 ("Q. What are the circumstances that would prompt you to mark a chat on the record? A. If I thought I was having a substantive conversation about a topic that was the subject of a litigation hold or that I received an instruction on a litigation hold.").

[115] *See* ███████, at 15:18-17:21, 83:24-84:3, 311:11-23 ("Q. Has anyone ever instructed you to delete chats? A. Never. Q. Have you ever actively deleted chats? A. No. Q. Has anyone ever instructed you to take conversations off e-mail into chat? A. No. Q. And you've been working with Google since 2012. Correct? A. Correct.").

[116] Hochstetler Report at ¶ 18.

HIGHLY CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

preservation obligations.[117] ██████, an ████████████████████ at Google, testified that Google did send regular reminders to Googlers on litigation hold regarding their obligations.[118]

64. Prof. Hochstetler's misunderstanding of the instructions issued to Google employees on litigation holds impacts the reliability of his opinions. At various points throughout his report, Prof. Hochstetler opines that Google "needed to [] preserve[]" or "should have [] retained" thousands of messages.[119] But those conclusions improperly assume, without basis, that all Googlers ignored the instructions that they were given.

iv.   Prof. Hochstetler's Analysis Does Not Account for the Custodians' Roles and Responsibilities

65. Prof. Hochstetler's reliance on the five custodians as a sample set from which to draw conclusions is even more unreliable given their roles and responsibilities, making any conclusions drawn from this sample likely irrelevant to the custodians not in his sample. Their work in Display Ads ranges from limited to non-existent in the 2022 timeframe that is the subject of Prof. Hochstetler's Report, and the Plaintiffs in this case took the deposition of only one of the five custodians. The evidence in this case also reflects that many of the agreed upon custodians left the Display Ads business long before 2022.[120] Custodians who have been deposed in this case have testified that, after leaving Google's Display Ads business, they did not use Chat to discuss their prior work in Display Ads.[121]

66. None of the five custodians work day-to-day in the Display Ads business. For example, Sundar Pichai became CEO in 2015.[122] As CEO, he is "not directly involved with [the Display Ads] business" and "clearly delegated responsibilities," relying on Display Ads executives "to update [him] if [he] needed to know something in more detail."[123]

---

[117]   *See* ECF No. 545, Ex. 5 at 4.
[118]   ████ Dep at 13:12-24. 230:17-231:9.
[119]   Hochstetler Report at ¶¶ 18, 81(b).
[120]   *See, for example*, GOOG-AT-MDL-C-000017436 at -438; *see also* GOOG-AT-MDL-C-000017436 at n.3.
[121]   Pappu Dep. at 37:14-19; 78:12-79-9 (stating she used Chat for "chitchat" and email for substantive communication and that, since leaving the display ads business in 2018, she has not been involved in "work related" discussions "with others who are working on DFP or AdX.").
[122]   Deposition of Sundar Pichai, August 30, 2024 (hereafter "Pichai Dep.") at 101:25-102:3.
[123]   Pichai Dep. at 18:25-19:6; 148:16-149:14.

HIGHLY CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

67. ███████ is ████████████████████ at Google.[124] Her role involves ██[125] As she explained in her pre-suit deposition, the transcript of which Plaintiffs have, she manages the small set of companies' relationships across various product areas, depending on the partner company.[126] I understand that Plaintiffs do not allege in their Complaint that █████ said or did anything in their Complaint and they did not depose her in this case.[127]

68. Dr. ████████ is ████████████████.[128] Prof. Hochstetler concedes that, for ██ ████, "there are only 36 sent messages in the Log Period, which is a much lower message volume than the other individuals represented in the Log Dataset."[129] That figure reflects an average of less than one Chat sent per day. Prof. Hochstetler speculates without foundation that the low messaging volume "could be due to the holiday season and does not necessarily reflect ████████'s typical message volume."[130] Based on that speculation, Prof. Hochstetler excludes ████████ as an outlier in drawing some of his conclusions.[131] I understand that the Plaintiffs do not allege in their Complaint that ████████ said or did anything and they did not depose him in this case.

69. ███████ was ████████████████████████████, as of September 2019, a role encompassing a wide range of products.[132] I understand that Plaintiffs do not allege in their Complaint that ████████ said or did anything and they did not depose him in this case.

---

[124] Deposition of █████, GOOG-AT-MDL-007170315 - GOOG-AT-MDL-007170598, July 28, 2020 (hereafter "████████.") at 12:9-21.
[125] ████████. at 13:22-25.
[126] ████████. at 14:3-25.
[127] Fourth Amended Complaint, *In Re: Google Digital Advertising Antitrust Litigation*, United States District Court, Southern District of New York, Civil Action No.: 1:21-md-03010-PKC, May 4, 2023 (hereafter "Complaint").
[128] ██
    https://people.ischool.berkeley.edu/~hal/bio.html.
[129] Hochstetler Report at ¶ 27.
[130] Hochstetler Report at ¶ 15, n.53.
[131] Hochstetler Report at ¶ 7(c), n.3 ("This statistic excludes information from Hal Varian, whose message logs represented only a small portion of the Log Dataset").
[132] GOOG-AT-MDL-002146534 at -535.

HIGHLY CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

70.　████████████ transitioned out of the ads business in October 2019.[133] I understand that Plaintiffs do not allege in their Complaint that ████████ said or did anything and they did not depose her in this case.

71.　The Log Dataset Prof. Hochstetler relies on suffers from significant quality issues, making it unreliable for extrapolation about the broader population. Drawing inferences from only a handful of employees that are non-representative of the target population (*i.e.*, 141 Google employees subject to a litigation hold related to this litigation at the beginning of 2022) poses significant risks of bias. The five custodians considered in Prof. Hochstetler's analyses are very unlikely to accurately reflect the range of behaviors observed across the entire body of employees, particularly given their specific roles within the company. I understand—based on the title of these employees and my understanding of the organization as informed by my review of the record—that these roles may inherently involve different levels and patterns of communication, leading to data that cannot be generalized to all custodians subject to a litigation hold in this case.

72.　Even Prof. Hochstetler himself admits that there exists "a wide variation in message volume between the different employees in the Log Dataset."[134] But rather than consult more reliable evidence in the record, such as sworn testimony and other produced documents reflecting information about these custodians' roles and responsibilities as they relate to the conduct at issue in this case, Prof. Hochstetler stops short. This failure to take into account available and reliable evidence in favor of extrapolating from a small and unreliable dataset infected by selection bias represents a methodologically unsound approach to the stated goals of Prof. Hochstetler's report.

---

[133]　*See* GOOG-DOJ-06548112 at -112, -114; GOOG-DOJ-11782550 at -522.
[134]　Hochstetler Report at ¶ 46.

HIGHLY CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

C.   **Prof. Hochstetler's Analysis Improperly Assumes Relevance Based on Selection Bias**

i.   *Prof. Hochstetler Relies on Biased Data, Rendering His Results Unreliable*

73. Prof. Hochstetler's opinion that there was a "loss of a significant number of relevant messages" based on "relevant chat conversations produced in this litigation,"[135] suffers from selection bias. As a result, any conclusion drawn from such a biased sample would not be representative of the broader population and any conclusions drawn from it are inherently unreliable.

74. Selection bias occurs when a sample is drawn in a manner that tends to exclude certain important or meaningfully distinct subgroups from the relevant population (i.e., the sample is drawn in a biased manner).[136] If selection bias is not addressed, reported results can suffer from systematic errors and thus not be generalized to the broader population at issue.[137]

75. I understand Prof. Hochstetler does not assert any form of expertise regarding the meaning of "relevan[ce]" from a legal standpoint. The focus of Prof. Hochstetler's "relevance" analysis is on two produced conversations involving Sundar Pichai. The logic underpinning Prof. Hochstetler's conclusion is that, because one portion of a conversation dated December 9, 2022, was produced, and because a second Chat conversation with the same <group_ID>[138] dated December 29, 2022, was also produced, then all other chat messages associated with the same <group_ID> would be "relevant" to this litigation.[139]

76. However, the produced Chat messages disprove this very point. I have examined the Log Dataset and determined that 12 "history on" messages were sent by Sundar Pichai with <group_ID> "AAAAbOkmVBc" from December 10 through December 28,

---

[135]   Hochstetler Report at ¶ 7(d).
[136]   2011 Kaye & Freedman, at p. 296. *See, also*, Tolga Bilgicer, *et al.*, "The Long-Term Effect of Multichannel Usage on Sales," *Customer Needs and Solutions* 2 (2015), pp. 41–56, at pp. 43–44.
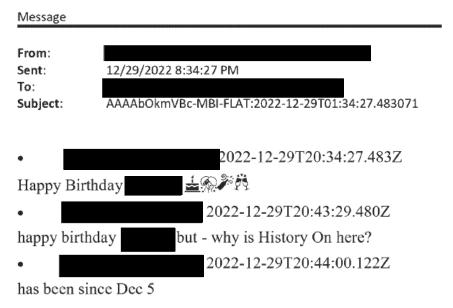[137]   2011 Kaye & Freedman, at pp. 222–223, 249.
[138]   Prof. Hochstetler purports to have analyzed sent messages in the Log Dataset using information from the <group_id> field. *See* Hochstetler Report at ¶ 30. However, Prof. Hochstetler purports to have "investigated" messages sent by Mr. Pichai in the group with ID "AAAAbOkmVBc" in the Log Dataset using the <space_id> field. *See* Hochstetler Report at ¶ 69.
[139]   Hochstetler Report at ¶ 7(d) and § VII.

HIGHLY CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

2022.[140] I understand from counsel that the produced chats do not contain any messages sent by the custodian during the same period. Therefore, I conclude that the messages sent by the custodian did not meet the production criteria in this case.

77. A new Chat conversation using the same <group_ID> began on December 29, 2022. This conversation appears to be personal in nature, and involves two Googlers wishing another Googler "happy birthday"[141]:

Exhibit 1

Chat Conversation in Group with ID "AAAAbOkmVBc" on Dec. 29, 2023[142]

78. I understand from counsel that the reason why a personal message of this nature would have been produced in this case is because it hits on the Discovery-on-Discovery Search Terms, in this instance, "History On." I understand that Google agreed with Plaintiffs in the Virginia Case to run searches for this term to avoid a discovery dispute about Chat retention, even though they were "unrelated to any party's claim or defense in th[at] action—so-called 'discovery on discovery.'"[143] I understand Google then

---

[140]   *See* Malkiewicz Workpapers.
[141]   GOOG-AT-MDL-007412395. *See* GOOG-AT-MDL-007412389 for Chat conversation in Group with ID "AAAAbOkmVBc" between Dec. 8, 2022, and Dec. 9, 2022.
[142]   GOOG-AT-MDL-007412395.
[143]   *See, for example,* May 26, 2023, Letter from Robert J. McCallum to Kelly D. Garcia at pp. 3–4.

HIGHLY CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

agreed to produce all of the output of discovery from that case in this case, regardless of whether they are relevant to the claims or defenses.[144]

79. This results in selection bias. Relying upon a document that hits solely on the search term "History On," Prof. Hochstetler has improperly defined "relevance" to include documents that were produced in a different litigation, and which were "unrelated to any party's claim or defense."

80. Prof. Hochstetler states that the goal of this portion of his analysis "was to identify logs corresponding to the produced conversations and determine whether any of them had Retention Settings toggled "off" at any point." Prof. Hochstetler asserts that "[t]he latter would indicate that not all messages relevant to this matter were preserved or made available."[145]

81. Prof. Hochstetler's reasoning is circular and conclusory. The Discovery-on-Discovery Search Terms in Schedule 3 target conversations in which a user toggled history off, among other such criteria, regardless of relevance to the case. And so the fact that a chat conversation is then produced does not, as Prof. Hochstetler asserts, "indicate that not all messages relevant to this matter were preserved or made available."[146] To the contrary, the impact of selection bias in this scenario means that Prof. Hochstetler is defining "relevance" by reference to the conclusion he seeks to draw. This is a flawed methodology that renders Prof. Hochstetler's opinion unreliable.

82. Starting from this flawed premise, Prof. Hochstetler then states that the Log Dataset reflects hundreds of additional messages sent in this group between December 29, 2022 and February 8, 2023 when history is forced on.[147] Prof. Hochstetler fails to note, however, that the Log Dataset reflects that an additional 14 messages were sent by Mr. Pichai in this group between February 8, 2023 and February 14, 2023.[148] These messages were retained but not produced because these chat messages did not meet production parameters in this case. This further undermines Prof. Hochstetler's

---

144 *See* May 26, 2023 Letter from Robert J. McCallum to Kelly D. Garcia at p. 4.
145 Hochstetler Report at ¶ 67.
146 Hochstetler Report at ¶ 67.
147 *See* Hochstetler Report at ¶ 77.
148 *See* Malkiewicz Workpapers.

HIGHLY CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

conclusion that other messages in the same group must have been "relevant to this litigation."[149]

83. Prof. Hochstetler is similarly wrong in concluding that four other chat "conversations, which are pertinent to this litigation, were generally impacted by the loss of message history."[150] Three of these conversations[151] were produced because they are responsive to the Discovery-on-Discovery Search terms in Schedule 3. As a result, they too are based on Prof. Hochstetler's erroneous definition of "relevance" and thus do not support the conclusion that Prof. Hochstetler seeks to draw.

84. The fourth conversation cited by Prof. Hochstetler also undermines the conclusion he seeks to draw. Prof. Hochstetler asserts that, with respect to <group_ID> AAAA-cp4gkQ, the Log Dataset reflects that ███████████ sent a single ephemeral message while sending six others that were permanently retained.

85. To locate the conversation referenced by Prof. Hochstetler, I reviewed the *Play* logs to identify a conversation in a <group_ID> in which ███████████ sends seven messages, one ephemeral and six permanently retained. I was able to identify a chat conversation that matches that description with the following log entries.

---

[149]   Hochstetler Report at ¶ 77.
[150]   Hochstetler Report at ¶ 80.
[151]   Hochstetler Report at ¶ 80.

HIGHLY CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

Exhibit 2

Messages Sent by ▮▮▮▮▮▮▮▮▮▮ **in Group with ID "AAAA-cp4gkQ"**

Time	<retention_status>	<group_id>
2023-01-19 13:27:00.400691 -0500 EST	EPHEMERAL_ONE_DAY	AAAA-cp4gkQ
2023-02-10 17:03:46.571473 -0500 EST	PERMANENT	AAAA-cp4gkQ
2023-02-10 17:09:26.697157 -0500 EST	PERMANENT	AAAA-cp4gkQ
2023-02-10 17:09:44.6155 -0500 EST	PERMANENT	AAAA-cp4gkQ
2023-02-10 17:10:20.403499 -0500 EST	PERMANENT	AAAA-cp4gkQ
2023-02-10 17:14:29.004883 -0500 EST	PERMANENT	AAAA-cp4gkQ
2023-02-10 17:20:20.214625 -0500 EST	PERMANENT	AAAA-cp4gkQ

Note: I ran Prof. Hochstetler's program "process_csvs.go" for <group_id> AAAA-cp4gkQ and filtered for backend actions CreateTopic and CreateMessage.

Source: Backup materials to the Expert Report of Prof. Hochstetler.

86. The above-referenced log entries correspond to a chat conversation that was preserved but not produced in this litigation because it did not meet production parameters, undermining Prof. Hochstetler's suggestion that all conversations sharing the same <group_ID> must be "relevant" to a litigation if any one of them was produced.

87. In light of the above, the more plausible explanation for the fact that history is switched off at various times is that the nature of the discussion did not pertain to matters relevant to this litigation. There is evidence in the record that custodians understood their obligations and that history could be switched off to have, for example, a personal conversation.[152]

88. Accordingly, Prof. Hochstetler's conclusion that "these conversations, which are pertinent to this litigation, were generally impacted by the loss of message history, meaning that relevant past messages were not retained or produced"[153] is wrong.

---

[152] *See, for example,* GOOG-AT-MDL-007431891 at -891 (discusses the February 8, 2023 retention change and stating "I'm glad not everyone is impacted, as they can then have non-work related conversations with each other without feeling their privacy has been violated"); GOOG-AT-MDL-011553781 at -786 ("Can turn off history for personal conversation"); GOOG-AT-MDL-007380107 ("I think history is off because sometimes we share pretty personal stuff here, and we don't have to think about it being recorded forever").

[153] Hochstetler Report at ¶ 80.

HIGHLY CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

ii.   *Prof. Hochstetler Conflates the Term "Conversation" with "Group"*

89. Prof. Hochstetler uses the term "conversation" inconsistently throughout his report. In many instances, Prof. Hochstetler uses the term "conversation" to refer to a chat conversation in the form it was produced in this litigation.[154]

90. However, in Section VI of his report, Prof. Hochstetler defines the term "conversation" as being synonymous with the term, "Group," opining that "Almost 90% of Conversations (Groups) For Each Individual Had The Chat Retention Setting Toggled Off At Some Point."[155] But in his Summary of Opinions, Prof. Hochstetler does not clarify that, by "chat conversations," he is in fact referring to "Groups" and not "produced conversations."[156]

91. This is misleading and confusing. By conflating "Groups" and "produced conversations," Prof. Hochstetler seeks to create the impression that Googlers frequently switch history on and off within the confines of produced conversations.

92. To compound that misleading impression, elsewhere in his report Prof. Hochstetler includes a "hypothetical conversation" in which two users do, in fact, switch history on and off repeatedly in the context of an eight-message exchange.[157] And despite this example being "hypothetical," Prof. Hochstetler asserts that it "demonstrates how changing the Retention Setting impacts message retention."[158] Accordingly, I view these assertions as misleading and speculative.

---

[154] *See, for example,* Hochstetler Report at ¶ 41 ("As I discuss in Section III above, the Retention Setting is applied per conversation. Therefore, if another participant changed the Retention Setting for a conversation that includes one of the individuals in the produced data, this change would apply to all participants of the conversation, including that individual.").

[155] Hochstetler Report at § VI.

[156] Hochstetler Report at § II.

[157] *See* Hochstetler Report at Figure 4 and n.83.

[158] Hochstetler Report at ¶ 22.

HIGHLY CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

D.   Prof. Hochstetler's Analysis Suffers From Various Other Flaws

> **i.   *Prof. Hochstetler's Analysis Does Not Account for Conversations for Which History Is Already Toggled On***

93. Prof. Hochstetler opines that "[n]one of the five individuals represented in the Log Dataset personally changed Retention Settings for Group Chats."[159] He bases this opinion on his analysis of the log action "DynamiteFrontendService-UpdateGroupRetentionSettings". This opinion lacks crucial context for several reasons.

94. As Prof. Hochstetler acknowledges, and as the logs show, a chat thread history setting can be updated by any participant in the thread, and the setting then applies to all participants on the thread.[160] Within the logs' <dynamite_server_extension> field, the following values represent the chat retention settings, *regardless of the user who enabled them*: "retention_state = PERMANENT" (history "on") or "retention_state = EPHEMERAL_ONE_DAY" (history "off"). Therefore, the observation that the individuals represented in the Log Dataset did not update this setting themselves does not imply that the history setting was turned off for all of their conversations. Any other participant in their conversations could have turned history on (i.e., enabled the "retention=permanent" setting) at any time, and this would not register in the logs as an action taken by the user even though it would have affected the user's conversation with the other participant.

95. Furthermore, conversations for which the retention setting was already set to "permanent" before the 68 day period covered by the logs would not show any participant changing the retention setting to permanent during the period covered by the logs. Such conversations would nevertheless all be retained, because they were set to history on at the beginning of the period covered by the logs, and no participant changed the history setting.

96. As Prof. Hochstetler neglects to state, his observation that no individual in the Log Dataset changed their history setting during the period covered by the Log Dataset

---

[159]   *See* Hochstetler Report at § V(B).
[160]   *See* Hochstetler Report at ¶ 41.

HIGHLY CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

necessarily implies that *none of the individuals represented in the Log Dataset ever turned history off during the period captured by the dataset.*

ii.   *Prof. Hochstetler's Analysis Relies on Unsupported Claims*

97. Prof. Hochstetler also makes unsubstantiated claims, resulting in his opinions being unreliable. For example, when estimating the number of messages allegedly lost in a year, Prof. Hochstetler assumes that half of the 141 custodians would leave Google in 2022.[161] However, Prof. Hochstetler provides no evidence in support of his assumption why half of the custodians would leave Google in 2022. Moreover, Prof. Hochstetler does not account for the fact that many Google employees left the Display Ads business before the timeframe covered by his Report.[162]

iii.   *Prof. Hochstetler's Report Misrepresents His Analysis by Improperly Excluding* ▮▮▮▮▮▮

98. When reporting the number of sent and received messages allegedly not retained by the custodians represented in the data, Prof. Hochstetler excludes information from ▮▮ ▮▮▮▮, presenting misleading and inflated findings.[163] He states that prior to February 8, 2023, "at least 92%" of sent and received chat messages by four custodians were not retained.[164] If Prof. Hochstetler had used his full sample by including ▮▮▮▮▮▮, his purported proportion of sent and received messages not retained by all five custodians would have been significantly lower.[165]

99. Prof. Hochstetler asserts that ▮▮▮▮▮▮'s lower message volume relative to the other four custodians in the data "could be due to the holiday season and does not necessarily reflect ▮▮▮▮▮▮'s typical message volume."[166] However, Prof. Hochstetler has no basis to assume that ▮▮▮▮▮▮'s message volume during this period differs from his regular message volume. The exclusion of information from ▮▮▮▮▮▮ inflates the

---

[161]   Hochstetler Report at n.4 and n.106.
[162]   *See* GOOG-AT-MDL-C-000017436 at n.3.
[163]   Hochstetler Report at n.3. Prof. Hochstetler writes that "[t]his statistic excludes information from ▮▮▮▮▮▮, whose message logs represented only a small portion of the Log Dataset, as detailed in Section V.D."
[164]   Hochstetler Report at ¶ 7(c).
[165]   Hochstetler Report at ¶ 7(c) and ¶ 48 (Table 9).
[166]   Hochstetler Report at n.53.

HIGHLY CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

statistic that Prof. Hochstetler reports concerning the number of sent and received messages allegedly not retained by the custodians in the data.

E.   Empirical Analysis Shows That Prof. Hochstetler's Analysis Is Undermined by Google's Chat Productions

100.   As noted above, on February 8, 2023, Google switched the default retention setting to "history on" for all Chat messages, and after that point, that setting could not be changed for any Chat messages involving employees who were subject to a litigation hold. The agreed upon discovery period in this case ended on March 27, 2023. And so it is possible to analyze the number of responsive chats in the 48 intervening days to test Prof. Hochstetler's assertion that "messages relevant to this litigation were lost."[167]

101.   Taking Sundar Pichai as an example, Prof. Hochstetler calculates that Mr. Pichai had 2,978 sent messages during the 68 days, averaging 66.18 messages sent per workday. As shown in Exhibit 3, extending this daily average over the 33 workdays in the post-February 8 period, one would expect 2,184 messages to have been sent during that timeframe. However, Google produced no Chat messages for this period, suggesting that none of the conversations met the established production criteria. While Prof. Hochstetler speculates that Mr. Pichai did not retain messages relevant to this litigation, an alternative explanation, consistent with his testimony, would be that he complied with the litigation holds, and none of his conversations during this period were relevant to the issues in this case.[168]

102.   Similar trends are observed for other custodians (see Exhibit 3). For instance, Prof. Hochstetler calculates that ▮▮▮▮▮▮▮▮▮▮ sent 5,230 chat messages during the Log Period, averaging 116.22 messages per workday. Extending this to the post-February 8 period yields an estimated 3,835 sent messages, yet only 153 sent messages were produced, implying a responsiveness rate of just 4.0%. Similarly, ▮▮▮▮▮▮▮ and ▮▮▮▮▮▮▮ had responsiveness rates of 9.5% and 5.6%, respectively, while ▮▮▮▮▮ had a responsiveness rate of 15.2%. These low implied responsiveness rates suggest

---

[167]   Hochstetler Report at ¶¶ 21, 77.

[168]   *See, generally,* Pichai Dep. at 63:24-64:2 ("You know, I typically did not conduct substantive business conversations, and I complied with my litigation holds.").

that only a small fraction of the chat messages sent during this period met the production criteria.

Exhibit 3

Responsiveness Rates Based on Prof. Hochstetler's Analysis of Messages Sent for Five Custodians

Custodian	[A] Number of Messages Sent During Log Period According to Prof. Hochstetler	[B] = [A] / 45 Avg. Number of Messages Sent During Log Period per Work Day	[C] = [B] * 33 Expected Number of Messages Sent per Work Day (2/8/2023–3/27/2023)	[D] Number of Produced Individual Messages Sent (2/8/2023-3/27/2023)	[E] = [D] / [C] Implied Responsiveness Rate (Sent Messages)
███	244	5.42	179	10	5.6%
	36	0.80	26	4	15.2%
	2,581	57.36	1,893	180	9.5%
	5,230	116.22	3,835	153	4.0%
Pichai	2,978	66.18	2,184	0	0.0%

Notes:
[1] Work days exclude weekends and U.S. observed holidays.
[2] Empty messages and messages that are undated are excluded.

Sources: Hochstetler Report, Table 7; GOOG-AT-MDL-008150322; GOOG-AT-MDL-008197494; GOOG-AT-MDL-008358061; GOOG-AT-MDL-008358076; GOOG-AT-MDL-008358088; GOOG-AT-MDL-008358090; GOOG-AT-MDL-009114686; GOOG-AT-MDL-009114771; GOOG-AT-MDL-009114814; GOOG-AT-MDL-009214611; GOOG-AT-MDL-011143288; GOOG-AT-MDL-012900020; GOOG-AT-MDL-014190290; GOOG-AT-MDL-014213973; GOOG-AT-MDL-018182202; GOOG-AT-MDL-018182208; GOOG-AT-MDL-018182220

103.   The limited number of produced messages compared to the expected total volume underscores the irrelevance of many Chats.[169] The low implied responsiveness rates support the conclusion that Prof. Hochstetler's assumptions about the retention and relevance of messages are overly broad, as the vast majority of messages logged during

---

[169] In Exhibit 3, to calculate the column D—the Number of Produced Individual Messages Sent Dated 2/8/2023 – 3/27/2023, inclusive—I requested Google's counsel to run a search for all Chat conversations produced in this litigation in this timeframe, for which Mr. Pichai, ███, ███, ███, or ███ are custodians. Google's counsel provided me with the following documents responsive to that search: GOOG-AT-MDL-008150322, GOOG-AT-MDL-008197494, GOOG-AT-MDL-008358061, GOOG-AT-MDL-008358076, GOOG-AT-MDL-008358088, GOOG-AT-MDL-008358090, GOOG-AT-MDL-009114686, GOOG-AT-MDL-009114771, GOOG-AT-MDL-009114814, GOOG-AT-MDL-009214611, GOOG-AT-MDL-011143288, GOOG-AT-MDL-012900020, GOOG-AT-MDL-014190290, GOOG-AT-MDL-014213973, GOOG-AT-MDL-018182202, GOOG-AT-MDL-018182208, GOOG-AT-MDL-018182220. I then supervised a manual count of the component messages within each produced Chat conversation. I excluded empty messages (i.e., messages with no content) from this count.

the analyzed period did not meet the criteria for production.[170] This further undermines his assumption that relevant messages were lost due to retention settings. To the extent these messages were produced as a result of the Discovery-on-Discovery Search Terms, even the numbers in Exhibit 3 overstate the relevance of these materials.

VII.   CONCLUSION

104.   For the aforementioned reasons, Prof. Hochstetler's primary opinion—"that the total number of messages not retained by Google employees subject to a litigation hold was a million and a half or more in 2022 only"—is unreliable because it is based on a series of errors and unsupported assertions.

Executed on November 26, 2024,

_____

Michal A. Malkiewicz

---

[170]   Notably, in his Table 9, Prof. Hochstetler misrepresents the "[n]umber (proportion) of logs for messages received that were not retained, full dataset." However, in his footnote 92, Prof. Hochstetler notes "there are typically (but not always) 3 log entries for received messages, so that the number of actual messages received is likely close to one third of the number of logs reported in columns 3 and 4 of this table." Hochstetler Report at n.92.